Pentium® III Processor Overview

- **Key elements of the Pentium® III Processor**
  - Streaming SIMD extensions
    - SIMD-FP
    - Prefetches
    - Streaming stores
    - New media instructions

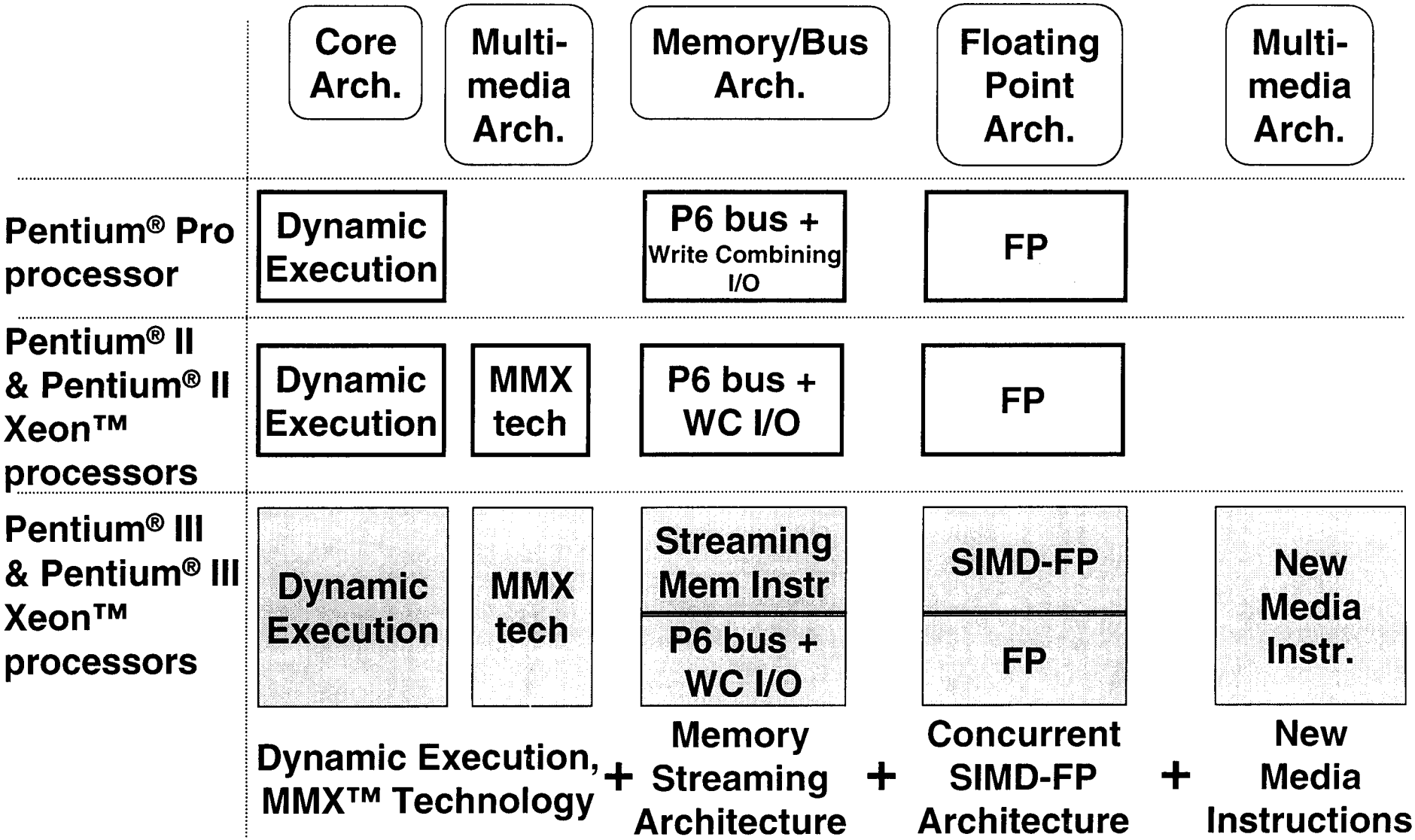
    } **70 New Instructions**
  - Processor Serial Number
- **Tuned for frequency scalability**
  - Circuit improvements for higher clock frequency
  - Memory streaming for higher bandwidth
- **Desktop and Pentium® III Xeon™ Processor workstation & server configurations**
- **5-layer metal 0.25 micron CMOS process technology**
  - 9.5M transistors

    ckaged using C4 technology

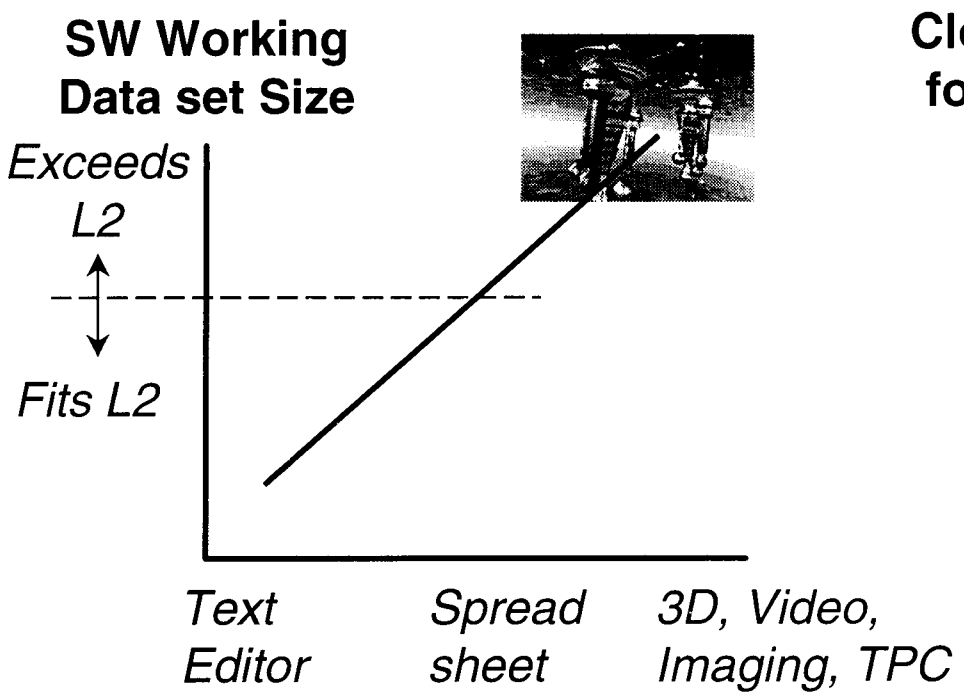
Outline

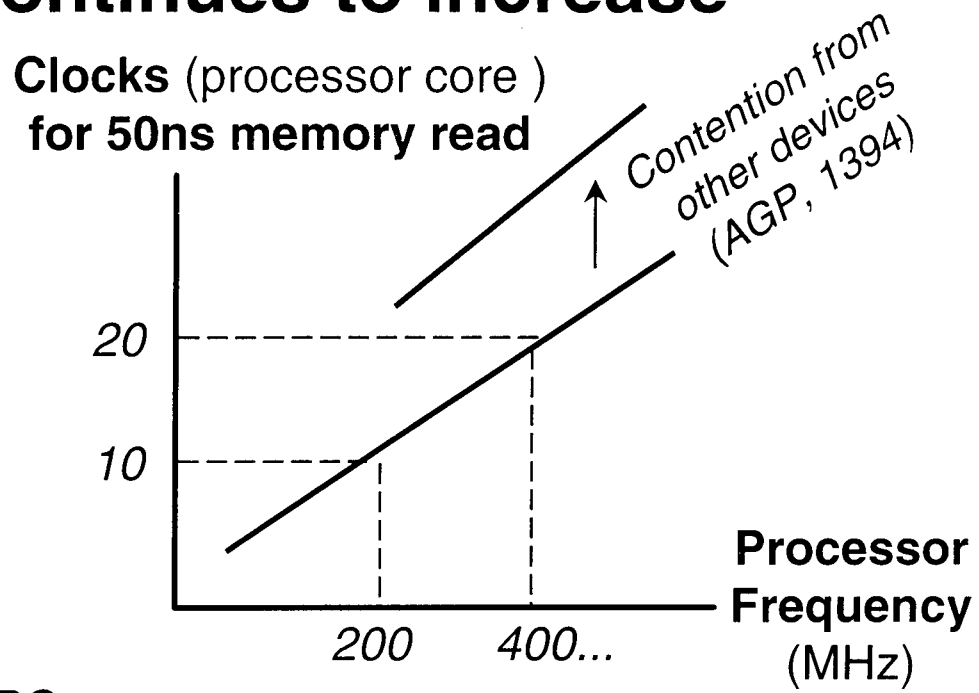
- Architecture overview
- Process & package technology overview
- Design & circuit implementation

Pentium® III Processor Evolution



Applications don't fit in Cache ...
Memory latency continues to increase





As working data sets >> L2...

❶ Memory & I/O Bandwidth needs ↑

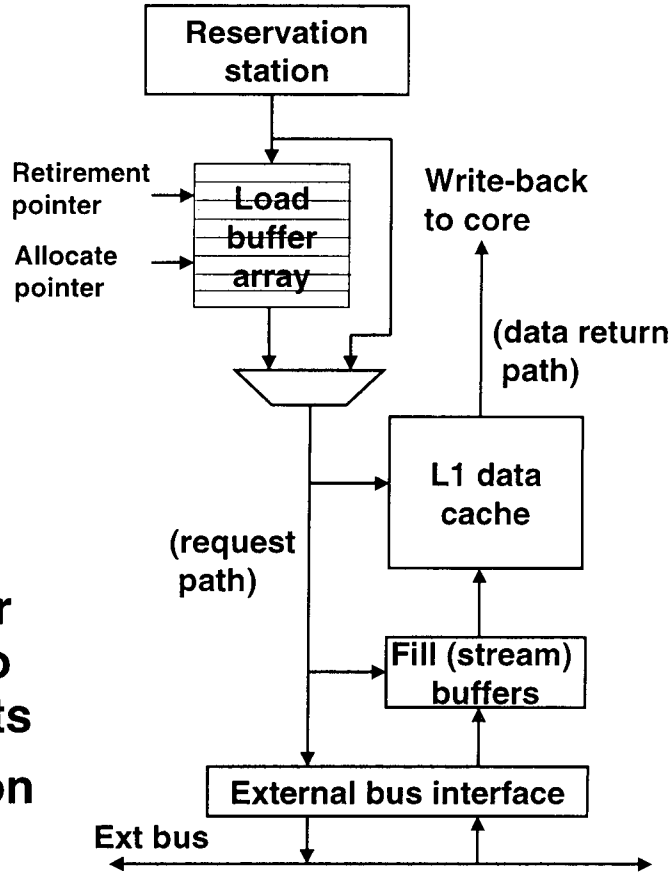
❷ Traditional cache heuristics not optimal (temporal locality, WB, RFO)

To make things worse...

❸ Memory is getting "further away" - relative latency ↑

'emory Latency & Bandwidth Limits Performance

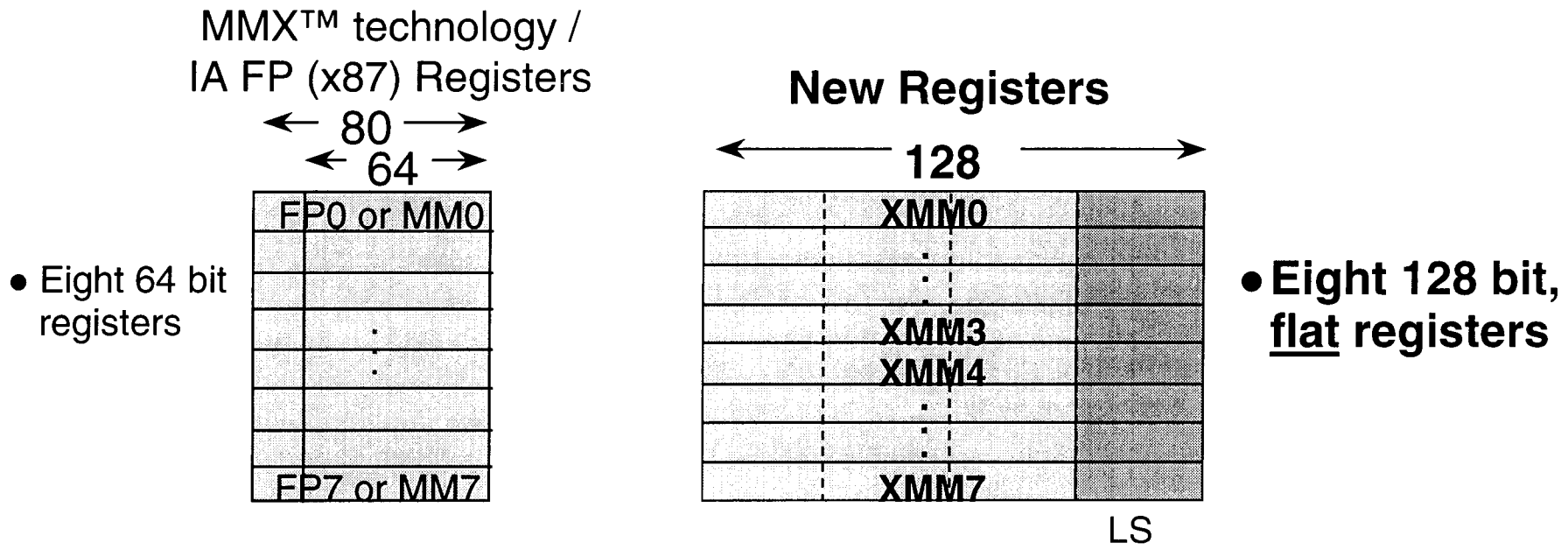
Memory Streaming
- Data prefetching
- Streaming stores



Description	Benefit
• Prefetch cache line into processor	Hides memory latency
• Prefetch operations non-blocking	Minimizes processor pipeline stalls due to external bus requests
• Store-through instructions write data to memory only	Avoid cache pollution for write-once data
• Partial stores combine to cache line size buffers before dispatch	Maximizes external bus throughput via bursts & pipelining of multiple requests

Improved bandwidth for data reads/writes

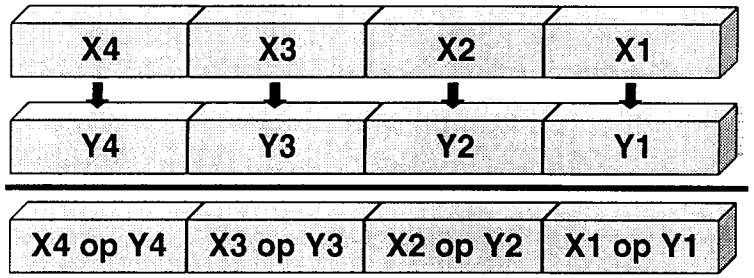
Concurrent SIMD-FP Architecture



- **Completely new separate state**
  - First time since 386
  - Concurrent SIMD-FP with MMX™ technology & IA-FP instructions
- **Conditional flow support**
- **Two modes: Flush-to-zero, IEEE-754**

Comprehensive set of SIMD-FP & SIMD-int Instructions

SIMD-FP instructions:



- Load, store
- Basic arithmetic: +, -, *, /, √
- Fast 1/x, 1/√(x)
- Logical, comparison
- "Swizzle" instructions
- Conversion: SIMD/scalar FP ↔ SIMD/scalar integer

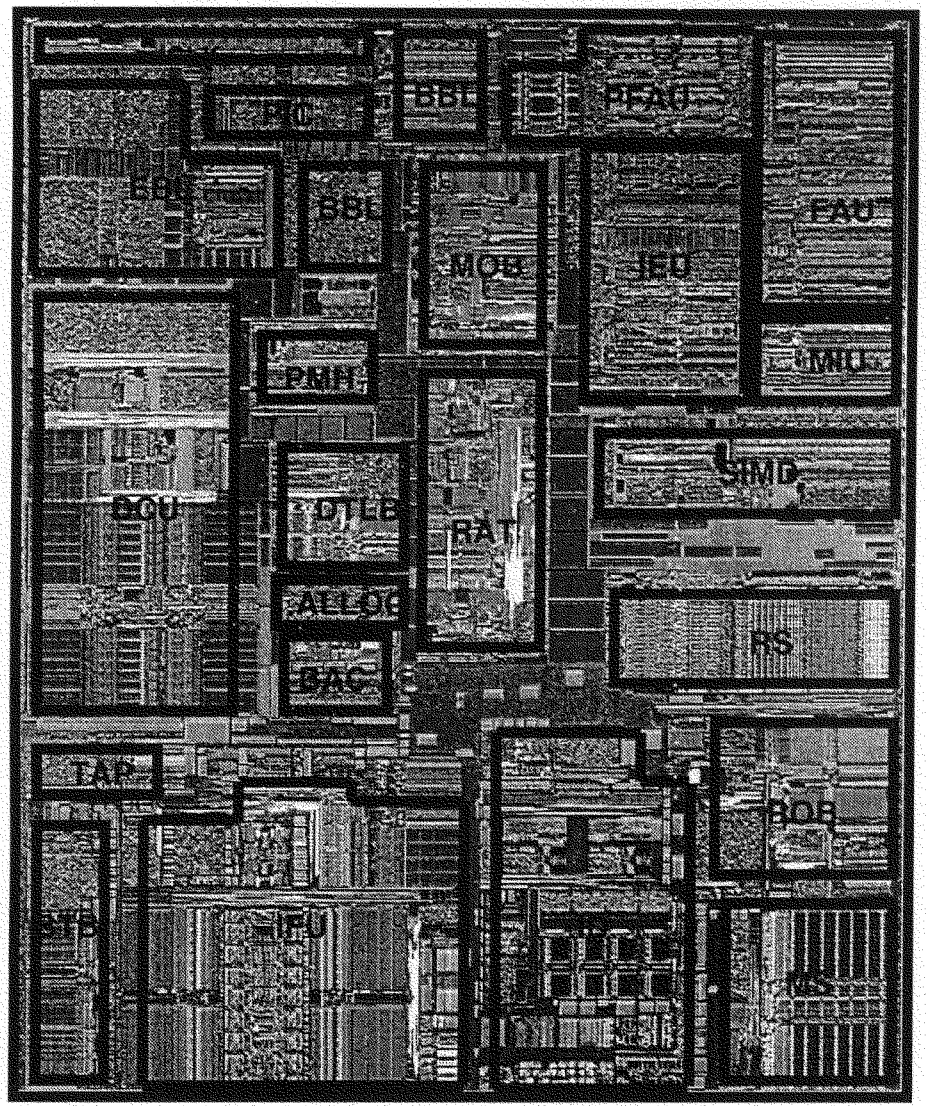
SIMD-integer instructions additions:

- Rounding average
- Sum of absolute differences
  , Max

Pentium® III Processor Benefits

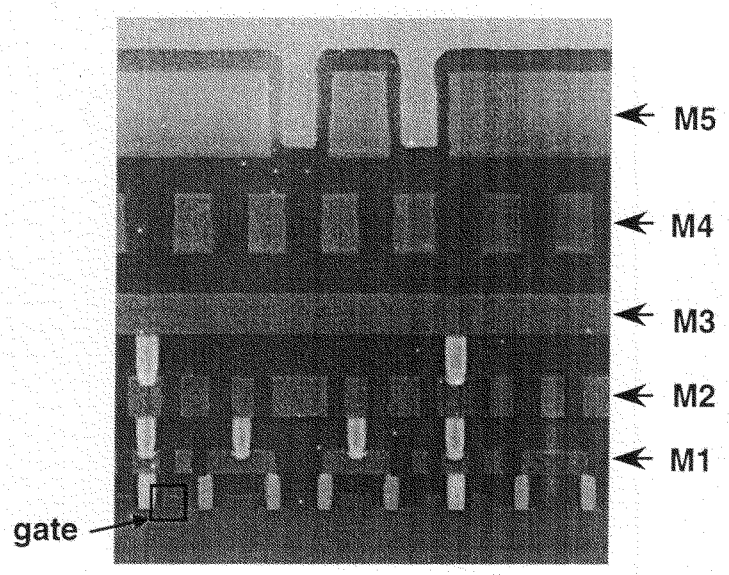
- **Real-time MPEG II full-resolution video & audio encode at 30 fps**
- **1.5 - 2x performance improvement for 3D transform & lighting kernels**
- **Large time reductions for speech recognition training**
- **Expect 5 - 20% gains for:**
  - TCP/IP checksum, Memory page zero/copy
  - Database Processing using PSE36
  - Digital Content Creation tools, Math Kernel Libraries
- **Processor serial number for manageability, security, and e-commerce**
  - Internet based applications

Process Technology



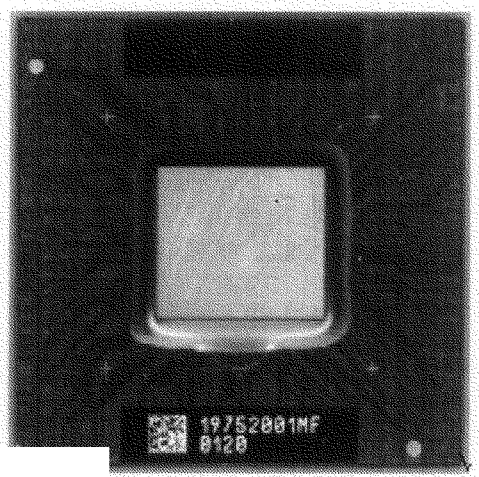
(after M1 processing)

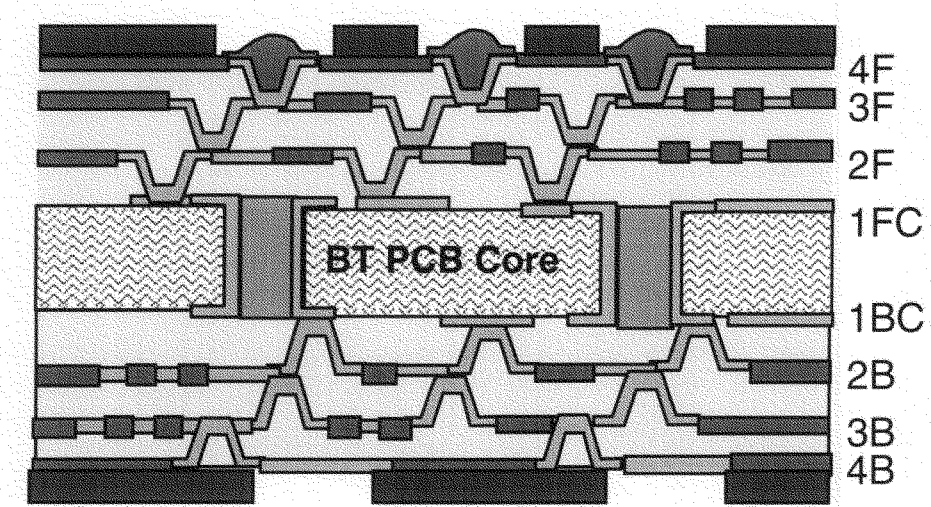
- Gate Oxide Thickness 40 A
- Gate Length 0.20 um
- Metal 1 pitch 0.61 um
- Metal 2 pitch 0.88 um
- Metal 3 pitch 0.88 um
- Metal 4 pitch 1.73 um
- Metal 5 pitch 2.43 um



C4 Packaging

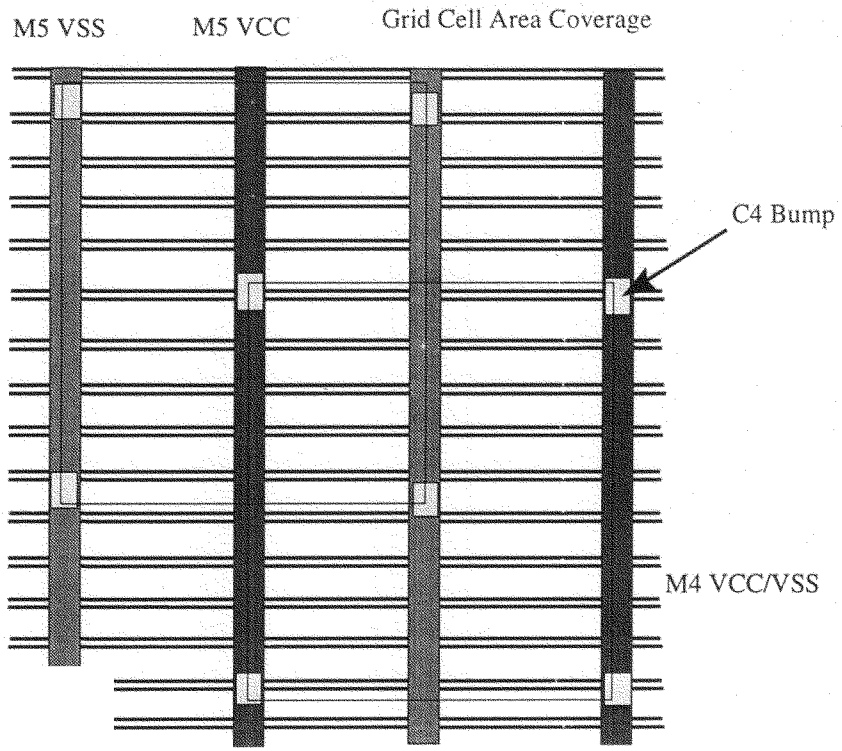
- **Six layer Organic Land Grid Array (OLGA) C4 package**
- **Two different C4 multi-grids for core logic and I/O circuits**
  - Core logic 252 um bump pitch, Periphery 235 um bump pitch
- **90 nF on-die decoupling capacitor**
- **Impedance & return paths optimized for I/O signal integrity**

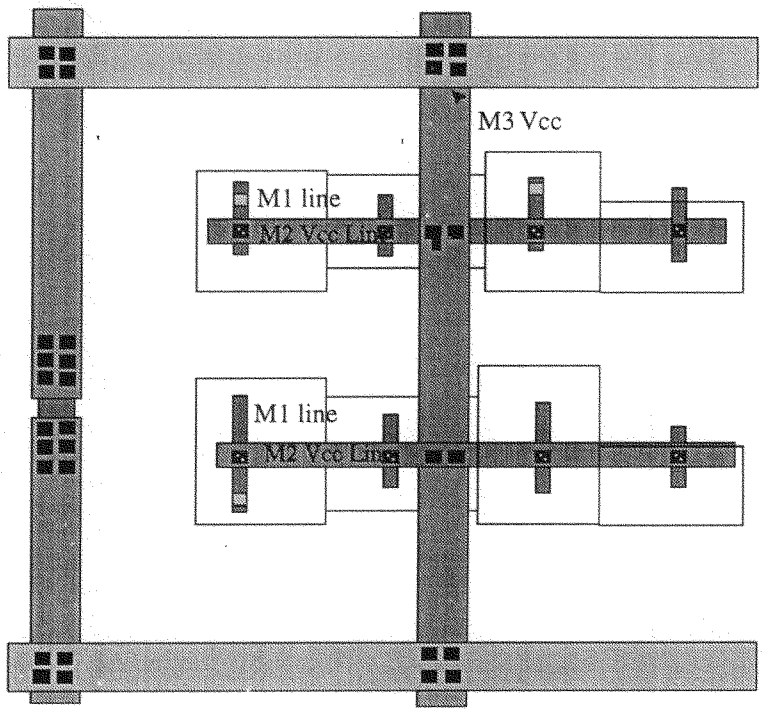




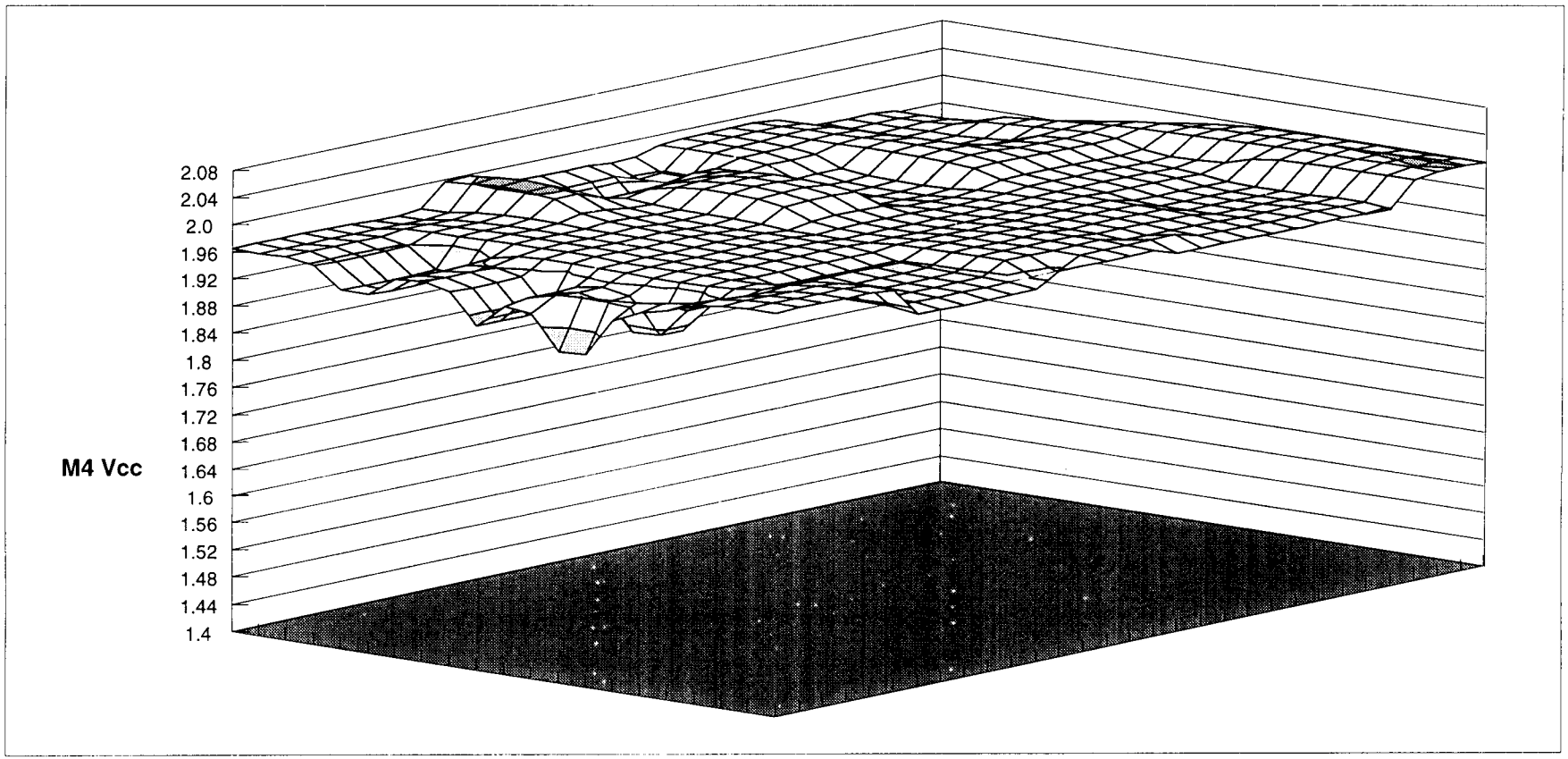
Power Delivery

- **Dedicated power & ground planes**
  - Two different Vcc supplies to enable low power applications
  - Package/cartridge level decoupling capacitors optimized for SSO minimization
- **Custom two layer power grids and localized tree distribution**

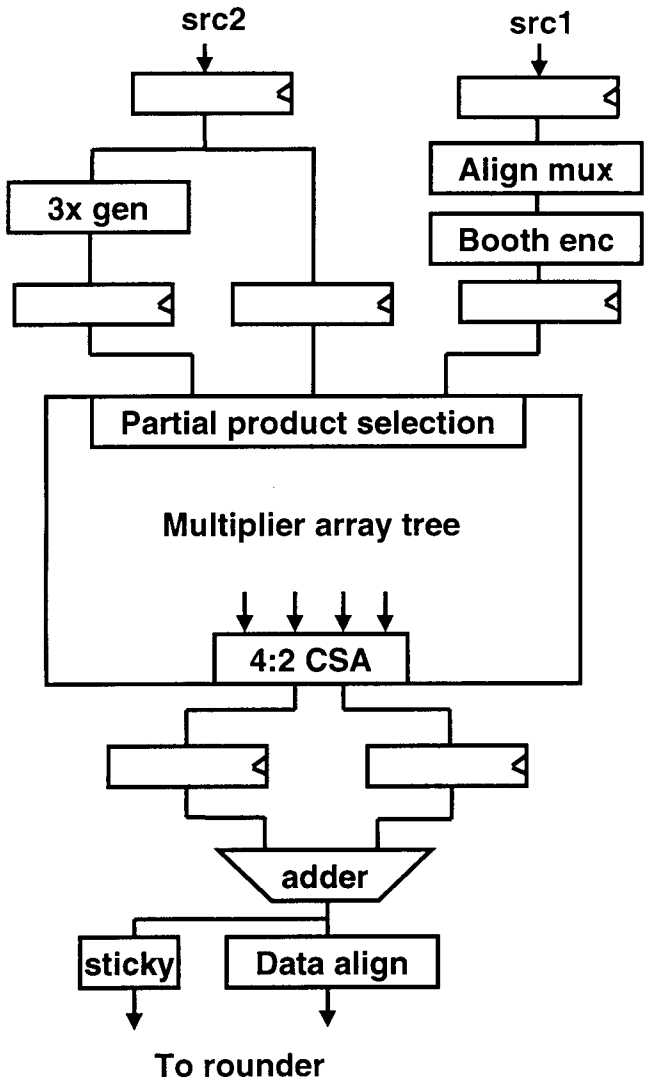




Global Power Grid Profile

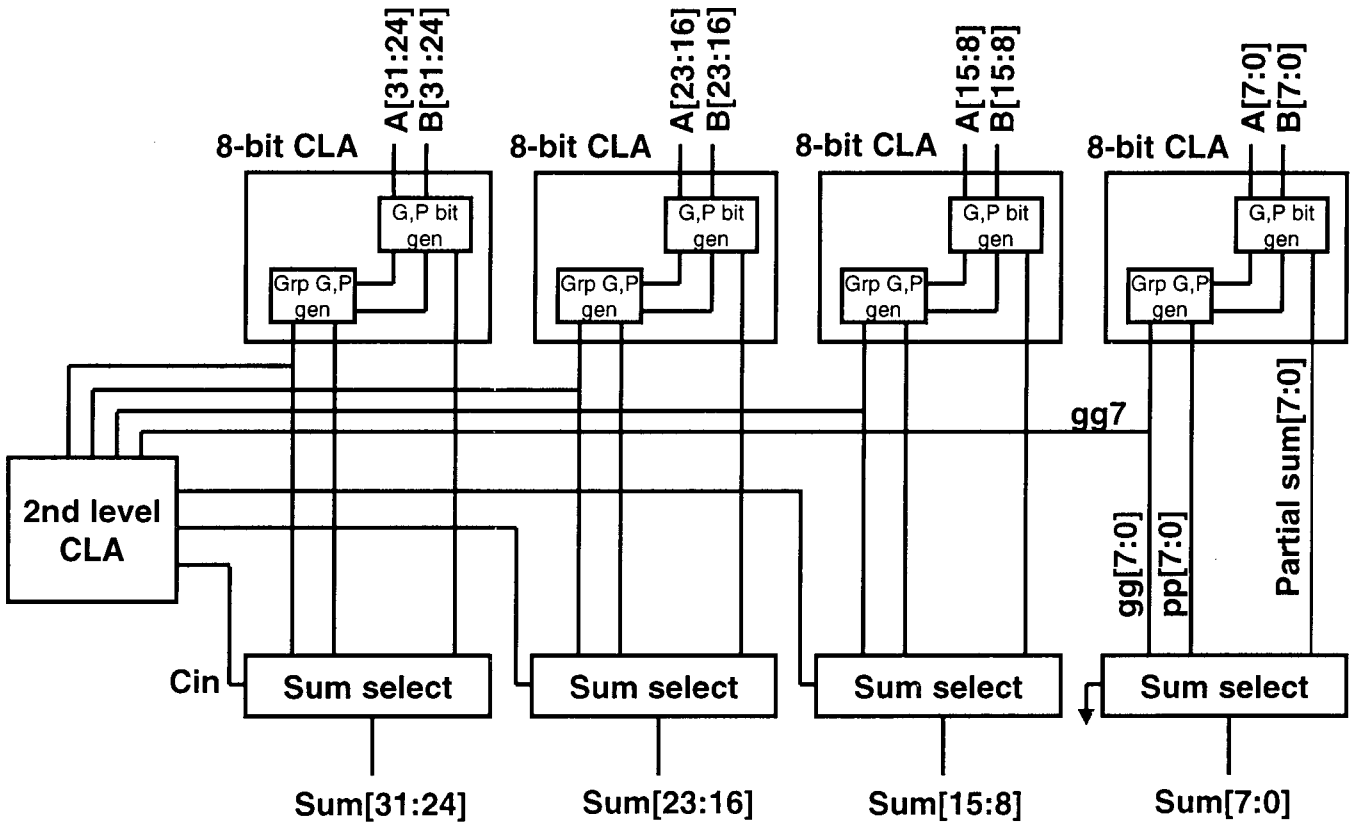


SIMD Floating Point Multiplier



- **Radix-8 modified booth encoding**
  - {0,+1,+2,+3,+4,-4,-3,-2,-1,-0}
- **Single pass multiplier array for SIMD-FP calculations**
  - enables fully pipelined operation
- **Zero insertion in multiplier array to avoid unwanted cross products**
- **Additional support for h/w exception handling**
  - pass-through on either source
  - rounder signaling for special number generation
  - Avoids lengthy microcode assists

SIMD FP Adder Mantissa Structure

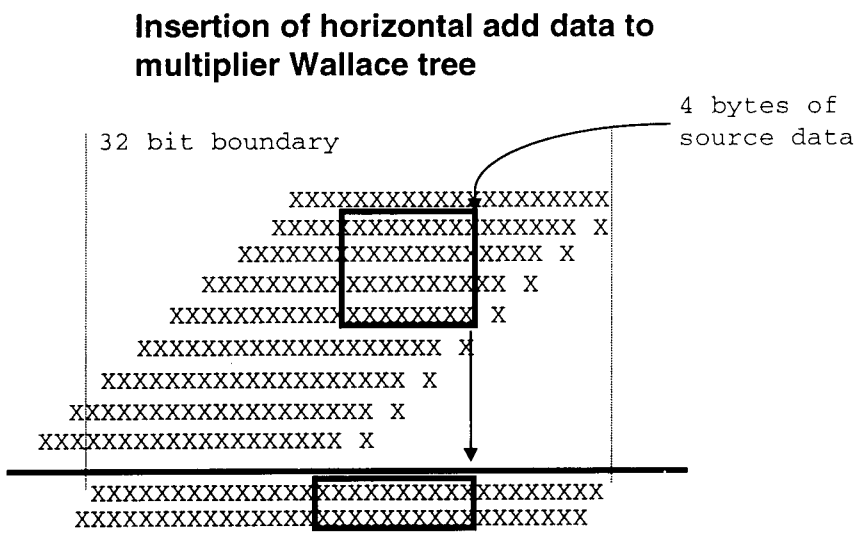


- 32-bit modified Kogge-Stone CLA adder
- Intermediate G, P bits and partial sums implemented using domino circuits
  - 1 clock phase for mantissa add function

SIMD-integer PSADBW (Sum of Absolute Differences) Implementation

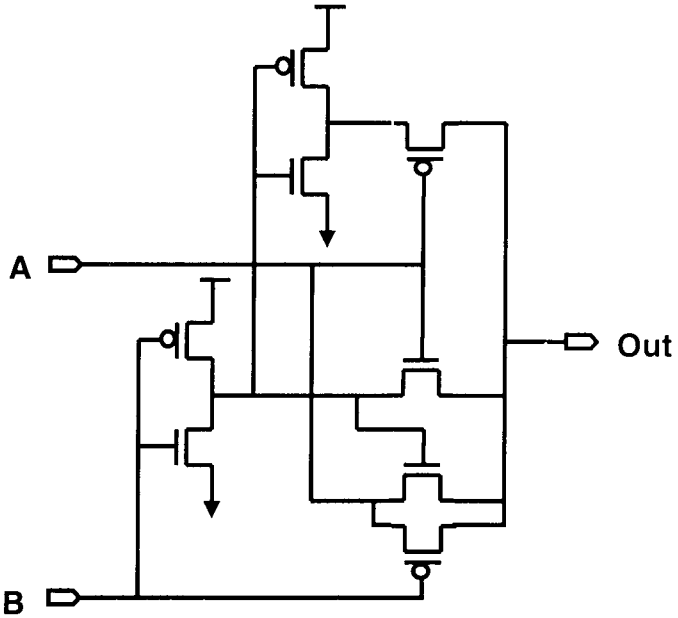
PSADBW functionality divided into 3 micro-operations:

1. Compute $(A_i - B_i)$ and $Carry_i$
2. Compute $|A_i - B_i|$ from $Carry_i$
3. Compute Sum $|A_i - B_i|_{i=1 \text{ to } 8}$

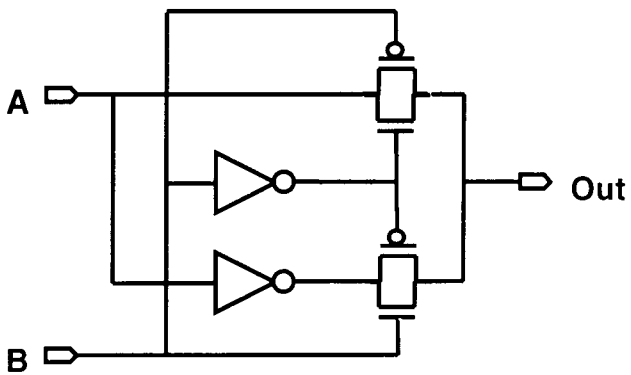


- **Powerful new MMX™ Technology instruction for video encode applications**
- **SIMD integer adders modified to write-back and source individual carry bits**
- **SIMD integer multiplier partial product selection logic retrofitted to perform horizontal add function**
- **Only increased SIMD integer unit by <2%**
  - Enables real-time MPEG II encode at 30 fps

Fast 8T XOR Gate



Improved 8T XOR gate



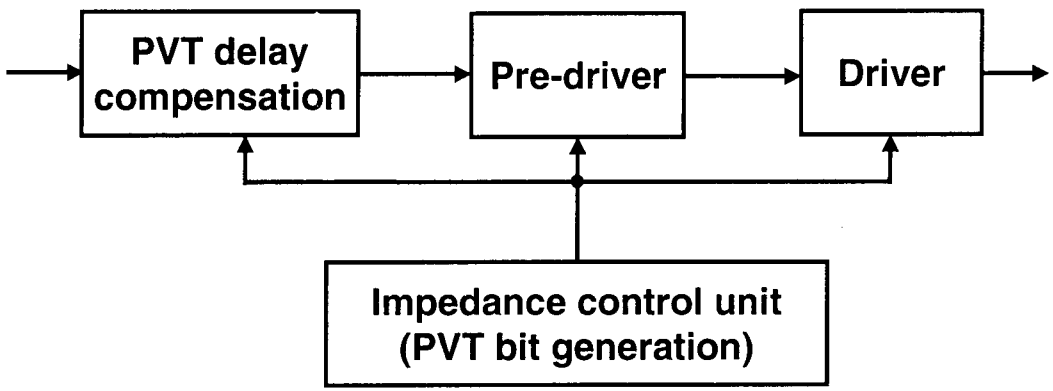
Previous version 8T XOR gate

- **Key building block for cache comparators**
  - Improve cache circuit timings without degrading self timed circuits.
  - Scales well with process shrink
- **Improved performance over previous 8T & 6T implementations**
  - 1 gate delay vs. 2 on the more critical input
  - Less contention compared to previous design
  - Improved back-writing and cross-capacitance impact
  - Up to 50% speed improvement

PVT Compensated I/O

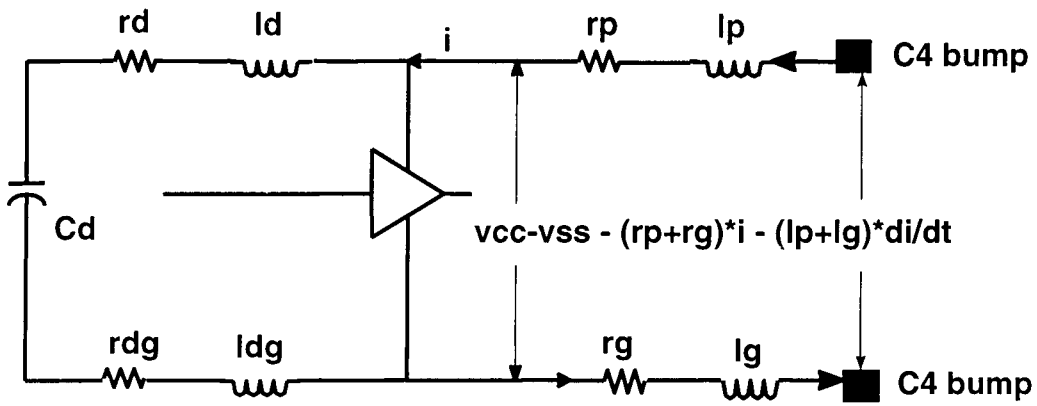
Challenges:

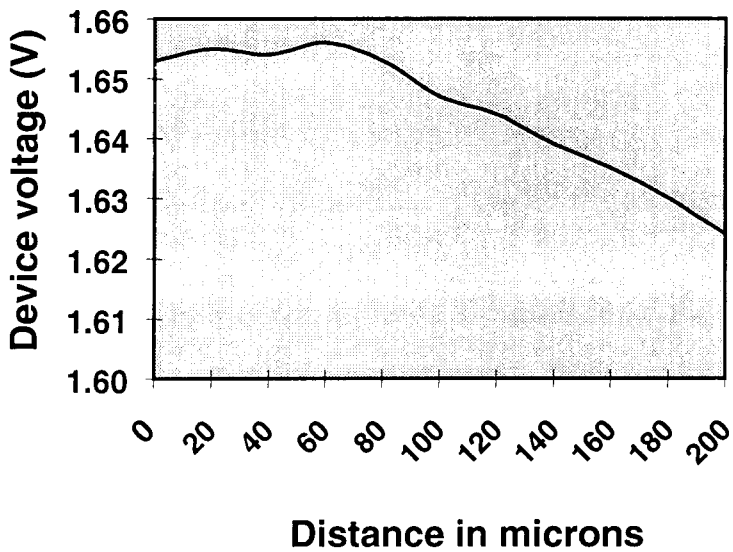
- **Source synchronous BSB Design.**
  - Data transfers at *half* and *full* core clock frequency
- **133/100/66 MHz common clock FSB with a single design**
  - Aggressive timing targets for both UP and MP topologies
- **I/O timings limited by Si due to process, voltage and temperature**



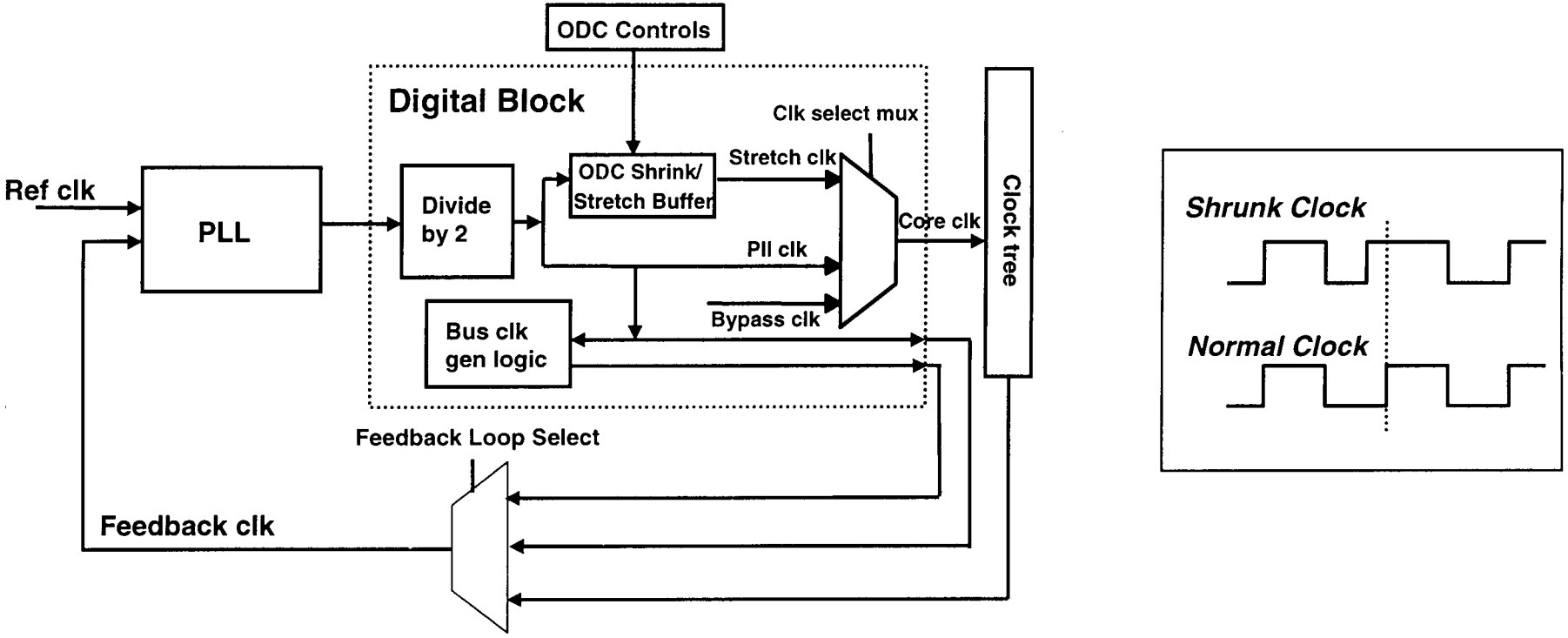
Current Starvation & Decoupling Caps

- Distributed lumped models used to determine placement strategy
  - Consideration for impact of 2 levels of neighboring decaps
  - Trade-off to minimize local and global voltage drops vs. silicon real estate
- Proximity key to effectiveness of de-coupling capacitors





On-Die Clock (ODC) Shrink

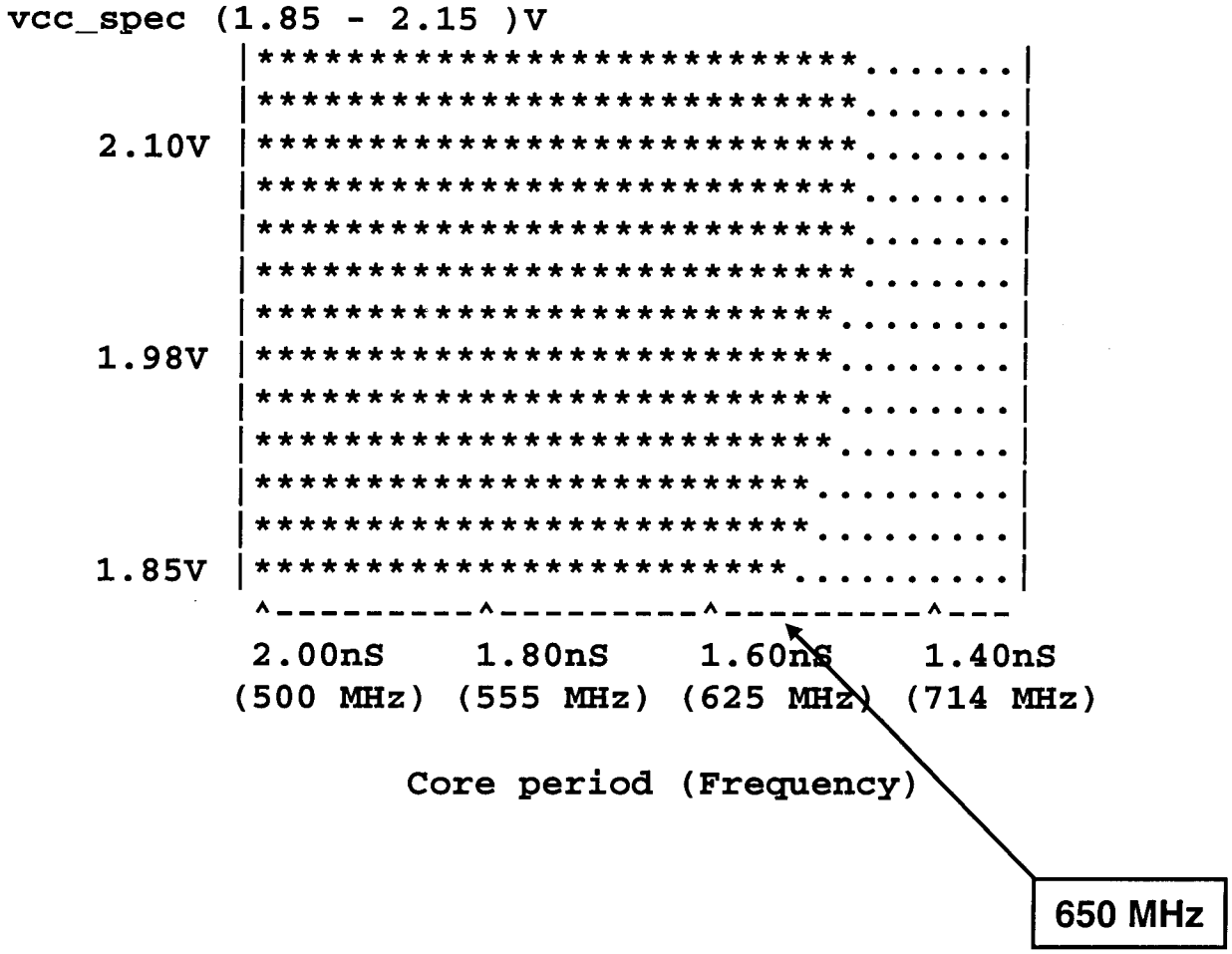


- **Allows *individual* cycle shrink & stretch ability for a specified phase of the global clock**
  - Avoids need for expensive test hardware.
- **ODC designed without any impact to conventional PLL**
- **Enabled rapid isolation & debug of on-chip speed paths**

Circuit Design Methodology

- **In-house tools used to increase productivity**
  - Parallel computation of full chip parasitic extraction
  - Incremental Fub level parasitic extraction
  - Automated speed path solver
  - Power grid IR checker
  - Xcap methodology
- **Correlated static timing database to post-silicon results**
  - Improved speed debug throughput

Clock frequency



Summary

- **New microarchitecture delivers scalable performance**
- **Comprehensive Streaming SIMD extensions to handle rich media data types**
  - New SIMD floating point data type
  - Data streaming
  - MMX™ Technology SIMD integer enhancements
- **Break through performance**
  - Real-time MPEG II video & audio encode at 30 fps
  - 1.5 - 2x performance improvement for 3D transform & lighting kernels
- **Currently operating up to 650 MHz**
  - Using today's 0.25 micron process technology